Journal of Science Research and Reviews

PRINT ISSN: 1595-9074
E-ISSN: 1595-8329
DOI: https://doi.org/10.70882/josrar.2025.v2i2.73
Homepage: https://josrar.esrgngr.org

Original Research Article

Machine Learning Algorithm for Optimal Yield Prediction of Cowpea (An IoT Smart Farming Approach)

***Benjamin T. Yecho, Oyenike M. Olanrewaju and Faith O. Echobu**

Department of Computer Science, Federal University Dutsin-Ma, Katsina State, Nigeria.
*Corresponding Author's email: ybenjamin@fudutsinma.edu.ng

KEYWORDS

Cowpea yield,
Crop yield prediction,
Machine Learning.

ABSTRACT

Cowpea (*Vigna unguiculata*) is a vital legume crop valued for its nutritional benefits and role in enhancing soil fertility; however, traditional farming practices often result in inconsistent yields due to environmental stresses and inefficiencies. This study explores how integrating Internet of Things (IoT), smart farming, and machine learning (ML) can optimize cowpea yield prediction and promote sustainable agriculture. The research focuses on implementing IoT-enabled smart farming systems with ML algorithms specifically Random Forest and AdaBoost to improve yield forecasting. IoT sensors were deployed to collect real-time data on critical parameters such as soil moisture, temperature, and nutrient levels, which were then used to train the predictive models. Performance evaluation using MAE, MSE, RMSE, and R^2 metrics revealed that Random Forest achieved perfect predictive accuracy (MAE, MSE, RMSE = 0.00; R^2 = 1.00), indicating strong generalization capability, while AdaBoost performed slightly less accurately (MAE = 0.05; MSE = 0.01; RMSE = 0.09; R^2 = 0.75), suggesting high accuracy but potential overfitting. The findings underscore the importance of soil nutrients and environmental variables in yield prediction and demonstrate that integrating IoT, smart farming, and ML particularly Random Forest holds great promise for advancing precision agriculture, increasing productivity, and fostering sustainable farming practices.

CITATION

Yecho, T. B., Olanrewaju, O. M., & Echobu, F. O. (2025). Machine Learning Algorithm for Optimal Yield Prediction of Cowpea (An IoT Smart Farming Approach). *Journal of Science Research and Reviews*, *2*(2), 131-139. https://doi.org/10.70882/josrar.2025.v2i2.73

INTRODUCTION

Cowpea (Vigna unguiculata) is a vital legume known for its nutritional value and nitrogen-fixing properties, but global yields remain low due to climate change, pests, and inefficient farming practices. Traditional methods struggle to address these challenges, highlighting the need for data-driven solutions like the Internet of Things (IoT), smart farming, and machine learning (ML).

The Internet of Things (IoT) refers to a network of interconnected physical devices embedded with sensors, software, and connectivity that enable them to collect and exchange data in real time. In agriculture, IoT systems allow continuous monitoring of environmental parameters such as soil moisture, temperature, and nutrient levels, enabling timely and data-driven farming interventions (Mishra et al., 2024).

Smart farming is the application of modern technologies such as IoT, data analytics, artificial intelligence (AI), and precision agriculture tools to improve decision-making, increase productivity, and reduce resource waste. It enables automated and optimized control of irrigation, fertilization, pest management, and crop monitoring, resulting in more sustainable and efficient farming systems (Alsayaydeh et al., 2024).

Machine learning is a subset of artificial intelligence that enables systems to learn from data, identify patterns, and make predictions or decisions without explicit programming. In agriculture, ML algorithms can analyze large datasets collected from sensors and other sources to predict crop yields, detect diseases, optimize input usage, and support resource management (Reshma et al., 2020). IoT sensors enable real-time monitoring of soil moisture, temperature, and nutrients, allowing precise interventions. ML algorithms analyze agricultural data to predict yields, detect diseases, and optimize resources, helping farmers make informed decisions that reduce waste and boost productivity.

Smart farming integrates IoT-based precision agriculture, continuously collecting data to enhance irrigation, fertilization, and pest control while minimizing environmental impacts. AI-driven automation further improves decision-making, replacing traditional trial-and-error approaches. Research underscores the economic benefits of these technologies, showing that IoT and ML improve yield predictions, optimize market planning, and increase profitability. Additionally, smart fertilizers and IoT-based monitoring significantly enhance cowpea yields. The integration of IoT, ML, and smart farming in cowpea cultivation offers a sustainable path to increased production, reduced resource waste, and improved food security. These innovations empower farmers to optimize growth conditions, mitigate risks, and ensure economic stability.

Machine learning (ML) algorithms play a crucial role in crop yield prediction, with gradient-based methods, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees, Random Forest, and Deep Learning techniques being widely used (Iliyasu, et al). These models leverage agricultural datasets to enhance predictive accuracy, optimize farming practices, and improve food security.

Gradient-based techniques, such as CatBoost, LightGBM, and XGBoost, have gained popularity for their efficiency in minimizing prediction errors. CatBoost, designed to handle categorical features effectively, has demonstrated the highest precision, achieving an accuracy of 99.123% (Mahesh & Soundrapandiyan, 2024). LightGBM, known for its speed and efficiency, uses histogram-based learning to reduce training time while maintaining competitive accuracy ($R^2 = 0.33$). XGBoost, valued for its scalability and ability to prevent overfitting through regularization, achieved an R^2 score of 0.31 in yield prediction. Comparative analyses indicate that CatBoost excels in accuracy, LightGBM is optimal for real-time applications, and XGBoost remains a robust choice for complex datasets (Shahhosseini et al., 2021).

KNN and SVM have been extensively studied for yield prediction based on soil and climate data. KNN predicts yield by comparing new data with past instances, making it effective for small datasets but computationally expensive for large ones. SVM, in contrast, performs well with high-dimensional data and captures complex relationships using kernel functions. While SVM is highly accurate, it requires careful tuning of hyperparameters such as kernel type and regularization.

Decision trees use hierarchical structures to segment datasets based on feature importance. However, they are prone to overfitting when used individually. Random Forest, an ensemble learning approach, mitigates this limitation by averaging multiple decision trees to enhance accuracy and reduce variance. Study like Pande, (2020) indicate that Random Forest outperforms other models, achieving 91.62% accuracy in yield prediction. Additionally, it provides feature importance rankings, identifying key yield influences such as soil moisture, temperature, and rainfall. In Nigerian agriculture, Decision Tree Regressor demonstrated a 72% accuracy rate (Shuaibu et al., 2024), reinforcing its relevance in yield prediction.

Koduri et al. (2019) developed a predictive model for crop production using AdaBoost regression, leveraging historical agricultural data from India. The study analyzed factors like rainfall, soil composition, temperature, and crop yield history. Data preprocessing included cleaning, normalization, and feature selection to improve accuracy. AdaBoost regression, using decision trees as weak learners, significantly enhanced prediction performance, achieving a high R-squared score, indicating reduced error and better generalization. However, the study lacked an extensive feature importance analysis, limiting interpretability for agricultural decision-making.

Chandraprabha & Dhanaraj (2023) integrated AdaBoost with Convolutional Neural Networks (CNN) and optimized performance using the Horse Herd Optimization Algorithm (HOA) to improve rice yield forecasting. The model used soil nutrient data and historical production statistics, achieving 95% accuracy, with precision and recall of 87% and 85%, respectively. The approach minimized error rates and optimized classifier weights. However, it did not incorporate climate factors like temperature and humidity, lacked real-time data integration, and did not assess computational efficiency for large-scale applications.

The study by Olanrewaju et al. (2024) presents an innovative approach to predicting maize yield by leveraging plant attributes and machine learning algorithms. Conducted at the Federal University Dutsin-Ma in Katsina State, Nigeria, the research focused on developing predictive models using Random Tree (RT), Random Forest (RF), and Artificial Neural Networks (ANN) to enhance agricultural planning and food security. The researchers collected maize yield data from experimental farms, emphasizing plant attributes as predictive features. They trained and evaluated three machine learning models RT, RF, and ANN using performance metrics such as Mean

Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE) to assess predictive accuracy. Among the models tested, the Random Tree algorithm outperformed the others, achieving: MAE: 0.093, RMSE: 0.096, RAE: 19.7%, RRSE: 19.2%.

Renju & Brunda (2024) employed a stacking ensemble learning approach integrating AdaBoost Regressor, Decision Tree, and Linear Regressor for crop yield prediction. Using Indian agricultural datasets, they optimized hyperparameters and handled missing data. Their ensemble model achieved an R^2 value of 98.92%, outperforming single-model techniques. AdaBoost Regressor contributed significantly to improving robustness. However, the study did not address scalability for broader agricultural applications.

Deep learning models, including Convolutional Neural Networks (CNNs) and Multilayer Perceptrons (MLPs), have shown promising results in yield estimation. CNNs, widely used in image processing, analyze satellite imagery to detect crop health variations, soil moisture levels, and growth patterns. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks excel in time-series forecasting, making them valuable for predicting seasonal yield fluctuations. Studies suggest that MLP and XGBoost perform exceptionally well in precision agriculture, enhancing decision-making through data-driven insights (Shawon et al., 2023).

Jiya et al, (2023)., in the research titled “Rice Yield Forecasting: A Comparative Analysis of Multiple Machine Learning Algorithms” gathered information on rice yield for the state of Katsina from 1970 to 2017 from the Nigeria Bureau of Statistics. Using logistic regression (LR), artificial neural networks (ANN), random forests (RF), random trees (RT), and naïve bayes (NB), models for predicting rice yield were developed using this dataset. The results showed that random forests and random trees fared better in yield prediction categorization tests. The TP rates of Naïve Bayes (NB), Neural Network (ANN), and Logistic Regression (LR) were 0.75, 0.19, and 0.75, respectively.

This study however explores how integrating Internet of Things (IoT), smart farming, and machine learning (ML) can optimize cowpea yield prediction and promote sustainable agriculture. The research focuses on implementing IoT-enabled smart farming systems with ML algorithms specifically Random Forest and AdaBoost to improve yield forecasting. IoT sensors were deployed to collect real-time data on critical parameters such as soil moisture, temperature, and nutrient levels, which were then used to train the predictive models.

MATERIALS AND METHODS

The methods adopted in this work include data collection, data preprocessing, choosing learning algorithm and training them to produce cowpea yield prediction model. Briefly, the diagram in Figure 1 captures the method conceptually.

Data source

The study was conducted at the Federal College of Agricultural Produce Technology, Kano, Nigeria, within the Northern Guinea Savannah ecological zone. The experimental site is located at an elevation of 686 meters above sea level, experiencing two distinct seasons: the dry season (November–April) and the rainy season (May–October). The region has a mono-modal rainfall pattern, with an annual mean precipitation of 1,110 mm, ranging between 950 mm and 1,270 mm. The average yearly temperature is approximately 25°C.

An experimental farm was established, consisting of plots of cowpea, each equipped with IoT sensors to monitor real-time environmental and soil conditions. The sensors measured soil nutrient levels (N, P, K), soil moisture, temperature, and climate factors such as wind speed, humidity, and rainfall. The structured dataset provided comprehensive records of soil conditions, climate variables, and cowpea growth parameters, ensuring accurate machine-learning analysis for yield prediction.

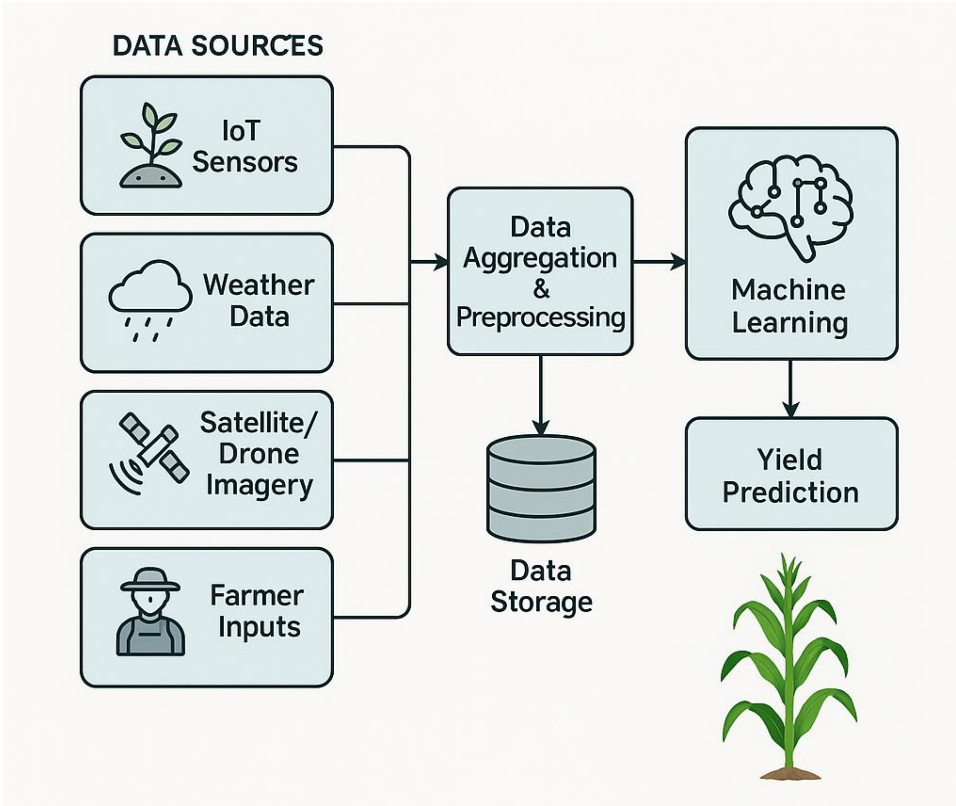

Figure 1: The conceptual model of yield forecasting

Feature selection

The features considered were Plant Height, Number of Leaves, Number of Flowers, Dry Weight (g), Fresh Weight (g), Average Temperature, Average Nitrogen, Average Phosphorus, Average potassium, Average Moisture and No of Pod(Yield).

Table 1: Statistical description of the data

	Plant height	Number of leaves	Number of flowers	Dry weight(g)	Fresh weight (g)	Average temp	Average N	Average P	Average K	Average moisture	Yield
count	40.000	40.000	40.000	40.000	40.000	40.000	40.000	40.000	40.000	40.000	40.000
mean	0.978	0.926	0.747	0.897	0.875	0.982	0.962	0.952	0.982	0.897	0.748
std	0.016	0.042	0.187	0.083	0.110	0.012	0.023	0.012	0.009	0.034	0.174
min	0.944	0.828	0.375	0.641	0.600	0.943	0.936	0.934	0.960	0.828	0.500
25%	0.965	0.897	0.625	0.872	0.800	0.975	0.946	0.947	0.978	0.884	0.571
50%	0.979	0.931	0.750	0.897	0.900	0.982	0.954	0.950	0.982	0.897	0.714
75%	0.995	0.966	1.000	0.974	1.000	0.994	0.984	0.955	0.988	0.905	0.875
max	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Data preparation

The collected data was pre-processed to remove any missing values, outliers, and unnecessary information. The mean, median, or mode technique were used to impute missing values, depending on the distribution of the data. The rationale behind removing outliers was that their existence would significantly affect the accuracy of the prediction model. The remaining data was used for further analysis. Data standardization was also done to ensure that each variable had similar ranges and distributions. 25% of the data was used for testing, while the remaining 75% was used for training. The number of pods per plot was used to calculate the yield of the plots. The maximum yield was 14 pods in a plot. The data was normalized by dividing the yield by 14. All other fields were normalized using the maximum value in the field. Normalization helps in scaling the features to a common range, which is crucial for models sensitive to the magnitude of input data. Table two contains IOT data collected.

Table 2: Data Collected

Plant height	Number of leaves	Number of flowers	Dry weight (g)	fresh weight(g)	Average Temp	Average N	Average P	Average k	Average moisture	yield
0.944	0.862	0.375	0.692	0.6	0.961	0.953	1.000	0.990	1.000	0.500
0.958	0.897	0.625	0.795	0.6	0.943	0.998	0.985	1.000	0.951	0.929
0.965	0.931	0.500	0.897	0.7	0.969	0.960	0.955	0.982	0.896	0.714
0.965	0.862	0.500	0.641	0.7	0.967	1.000	0.969	0.960	0.828	0.643
1.000	0.966	1.000	0.974	1.0	0.994	0.936	0.950	0.982	0.905	1.000

Model development environment

The models were implemented using python programming. Python package and libraries such as Scikit-learn, Pandas, NumPy, Matplotlib & Seaborn were used. Along with other features for transforming data into the right format for mining, the software includes several tools for data pretreatment and visualization.

Prediction Algorithms

Two machine learning algorithms were used for yield prediction: random forest and adaboost. These classifiers have their strength and weakness.

RF is a form of learning algorithm that generates a tress based on the attributes from the dataset, where each tree is itself a classification tree. Several random samples are generated from which randomized trees are developed. It initially randomly samples the complete data set, following which many decision trees are generated. Each tree is trained using a random sample from which it was built. All of the decision trees' predictions are then combined into a single tree for a single output. If multiple trees are trained and a greater number of them predict that an object belongs to class Y, and one says no, the final random forest prediction will be class Y.

AdaBoost (Adaptive Boosting) is a machine learning algorithm that enhances the performance of weak classifiers by combining them into a strong classifier. It works by training multiple weak classifiers sequentially, where each new classifier focuses more on the samples that previous classifiers misclassified. The final model is a weighted sum of these weak classifiers (Cao et al., 2013).

The AdaBoost algorithm follows these steps: Assign equal weights to all training samples.

The algorithm trains a weak classifier on the data, Compute the error of the classifier and increase the weights of misclassified samples so that the next classifier focuses on them. It follows the produce repeatedly until a number of iterations are reached. As the final stage, it combines the weak classifiers into a final strong classifier using a weighted majority vote (Charlin, 2004).

Model Evaluation

A good number of metrics methods exist to measure the accuracy of machine learning model. The metric used in this work that measure the error rate of the model include RMSE, MSE, MAE and R-squared (R^2) Score. Mathematical equations of the matrices are given in the following equations 1 to 4.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (1)$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (2)$$

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (4)$$

RESULTS AND DISCUSSION

Figure 2 presents the graph of the yield in all the plots. Figure 3 shows the Adaboost model of the yield prediction while figure 4 shows the Random Forest model of the yield prediction. The AdaBoost model was trained using a stump as the base learner with 70% of the dataset used for training and 30% for testing.

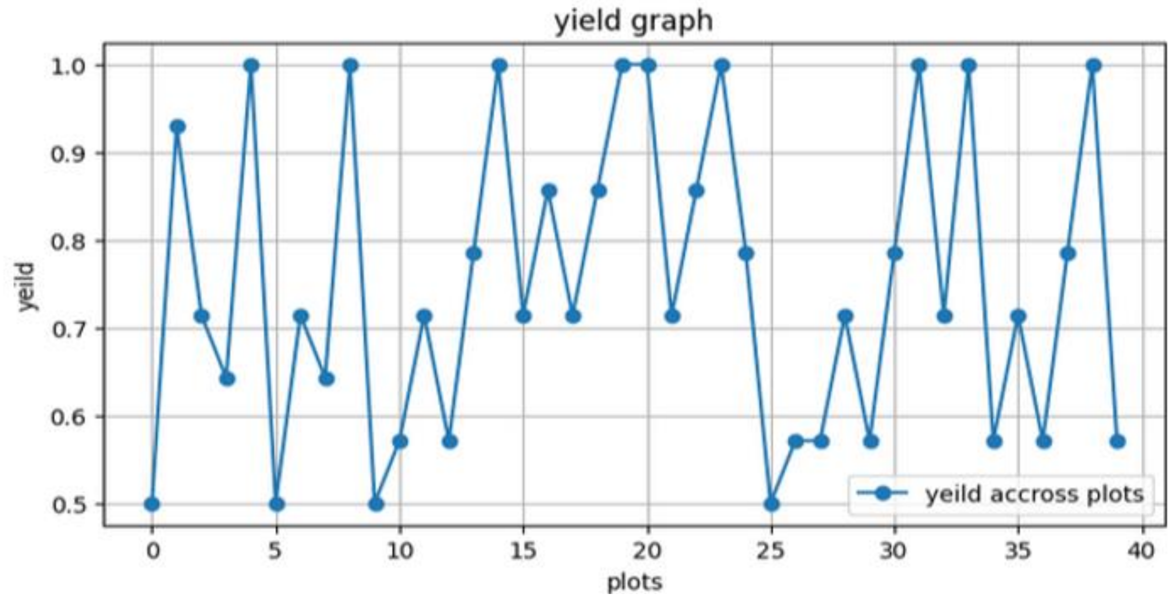

Figure 2: Yield of all the plots in the farms

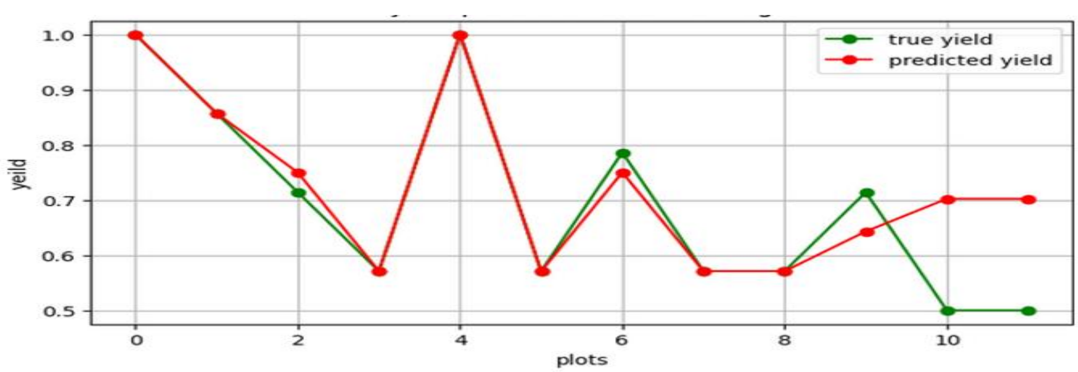

Figure 3: Adaboost model for prediction of cowpea

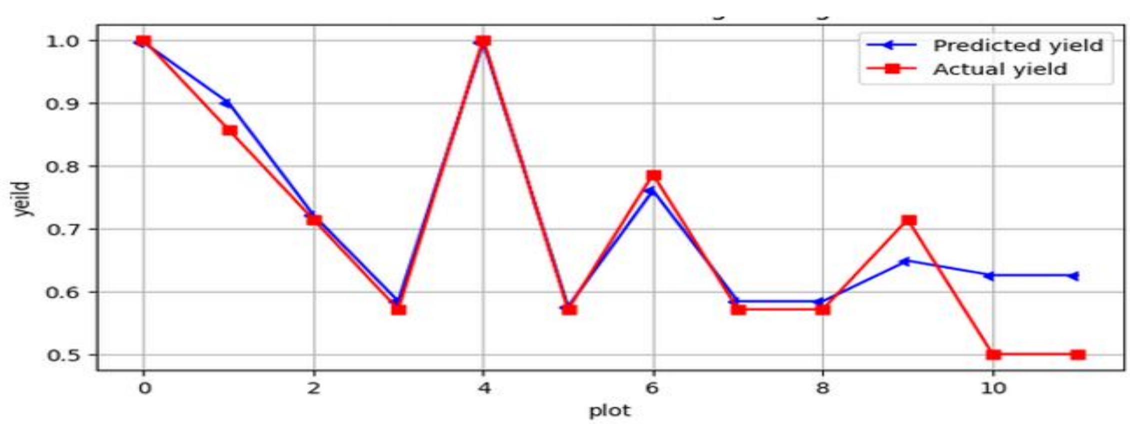

Figure 4: The Random Forest model of cowpea yield forecasting

Figure 2 represents the plot of all cowpea yield from all the plots. In the said figure, the yield was not uniform across the various plots, though the same fertilizer and treatment were provided. The result showed that out of 40 records of the yield, about 11 of the plots produced maximum yield of 1 (14 pods), 4 plots produced moderate yield (half of the maximum yield) while the majority were between 0.6 to 0.9. This shows the impact of the IOT in application of needed nutrients. The employment of IOT in application of needed fertilizer based on the needs of the plant resulted in many of the plots yielding very high.

Table 3: Yield Data

Pot	Plant Height	Number of Leaves	Number of Flowers	Dry Weight (g)	Fresh Weight (g)	Average Temp	Average N	Average P	Average K	Average Moisture	No of Pod(Yield)
1	13.4	25	3	2.7	9.7	31	17.39	23.27	54.11	38.6	7
2	13.6	26	5	3.1	10.5	30.42	18.22	22.92	54.63	36.71	13
3	13.7	27	4	3.5	11.9	31.27	17.52	22.23	53.67	34.57	10
4	13.7	25	4	2.5	11.3	31.21	18.25	22.54	52.46	31.98	9
5	13.7	26	5	3.4	12.6	30.38	17.65	22.48	53.58	35.38	9
6	13.7	24	6	3.1	12.8	31.42	18.03	22.29	53.18	35.02	7
7	13.7	26	5	3.5	13.2	32.12	17.22	21.81	52.5	36.48	10
8	13.7	25	6	3	13.5	32.26	17.36	22.23	53.25	35.57	9
9	13.7	26	6	3.1	14.1	31.74	18.17	22.14	53.91	32.08	11
10	13.9	27	5	3.4	15.1	31.43	17.41	22.13	53.69	32.93	7
11	13.7	26	5	3.5	13.1	31.67	17.26	22.03	54.01	34.51	8
12	13.7	28	6	3.9	12.8	31.12	17.69	22.33	53.44	33.4	10
13	13.7	27	5	3.6	14.4	31.9	17.35	22.15	53.59	34.63	8
14	13.9	26	4	3.5	14.7	31.62	17.95	21.73	53.07	34.48	11
15	14.1	29	5	3.4	14.4	31.81	18.22	21.98	54.46	34.67	13
16	13.9	26	6	3.4	15.6	31.49	18.16	21.89	54.55	34.14	10
17	14.1	29	9	3.2	16.2	32.08	17.86	22.15	53.85	34.89	12
18	14.2	26	7	3.4	15.7	31.45	17.68	22.33	53.25	38.39	10
19	14.1	29	7	3.9	16.9	32.18	18.21	22.47	53.63	32.54	12
20	14.2	28	8	3.8	16.8	32.07	17.08	22.11	53.66	34.94	14

Figure 3 shows the result of Adaboost model predicting the yield of cowpea based on fertilizer applicable. The model was able to track the dynamic changes in the yield across the plots. The model had minor errors in the third data and the last 2 values. The same result was observed in figure 4, where the model of yield prediction using Random Forest produced an excellent result similar to the Adaboost, with lower error observed in the last two data. The 2 models were able to track the actual yield across farms in all the plots, with slight prediction errors in predicted values surpassing actual yield in the last 2 plots, though the error of random forest is lower than the Adaboost. The accuracy of each model is presented in table 4.

Table 4: Model accuracy

	MAE	MSE	RMSE	R^2
Random Forest	0.00	0.00	0.00	1.00
AdaBoost	0.05	0.01	0.09	0.75

In table 4, the result of accuracy measure of the models in terms of MAE, MSE, RMSE, and R^2 were 0.00, 0.00, 0.00, 1.00 and 0.05, 0.01, 0.09, 0.75 for Random Forest and AdaBoost respectively.

Discussion

This study examines the performance of two distinct models for predicting cowpea yield based on climatic variables. The models were evaluated using error-based metrics to assess their accuracy in predicting numerical values. The developed models were each evaluated based on performance metrics of MAE, MSE, RMAE, and R2. From the analysis of the metrics, Random Forest demonstrated better performance than the other model results by achieving the lowest error rate. The error rate was all 0 when calculated using 2 decimal figures while the error rate of Adaboost was 0.05 for MAE.

Random Forest achieved perfect predictive accuracy (R^2 = 1.00) compared Adaboost R^2 of 0.75. this performance Random Forest can be due to its ability to handle complex relationships within the dataset. AdaBoost performed well but exhibited slight overfitting tendencies, particularly for outlier samples. These findings highlight the effectiveness of ensemble learning in agricultural yield prediction and emphasize the importance of feature selection and model tuning for optimal results.

The result of the two models suggests that both models were able to predict cowpea yield based on fertilizer application with an acceptable level of accuracy.

This study highlights the potential of machine learning in precision agriculture, particularly for cowpea yield prediction, by applying Random Forest and AdaBoost algorithms to IoT-generated climatic data. Random Forest emerged as the more accurate and stable model, outperforming AdaBoost, though both demonstrated usefulness in predictive tasks. The integration of smart farming technologies supports improved decision-making, efficient resource management, and increased agricultural productivity. The developed models are recommended for use in Nigeria to assist government planning and enhance food security. Future research should focus on expanding datasets, incorporating real-time data, optimizing hyperparameters, evaluating economic feasibility for large-scale deployment, and exploring hybrid models to further improve predictive performance. our study centers on cowpea yield prediction using climatic variables and evaluates Random Forest (RF) and AdaBoost models. It found that RF outperformed AdaBoost, achieving perfect predictive accuracy (R^2 = 1.00) and zero error (MAE, MSE, RMSE = 0.00) when rounded to two decimal places, while AdaBoost showed a MAE of 0.05 and R^2 of 0.75, with slight overfitting on outliers. This suggests RF's robustness in handling complex relationships in climatic data and supports ensemble learning as effective for agricultural prediction.

In contrast, the study by Olanrewaju et al. (2024) focuses on maize yield prediction based on plant attributes collected from experimental farms in Nigeria. It compares Random Tree (RT), Random Forest (RF), and Artificial Neural Networks (ANN). Here, Random Tree achieved the best performance, with MAE = 0.093 and RMSE = 0.096, indicating its superiority over RF and ANN in that context. The metrics used MAE, RMSE, RAE, and RRSE suggest moderate predictive accuracy, emphasizing the value of plant-based features.

CONCLUSION

This study demonstrates the potential of machine learning in precision agriculture, specifically for cowpea yield prediction. Random Forest and AdaBoost models were applied to IoT-generated data, with Random Forest emerging as the superior model. The results support the integration of smart farming technologies to enhance decision-making, improve resource management, and increase agricultural productivity. Future research should focus on expanding datasets, incorporating real-time data, and evaluating economic feasibility for large-scale adoption. This paper developed cowpea yield prediction model, using climatic data. The work used two (2) machine learning algorithms: Random Forest (RF) and Adaboost. We therefore recommend that our models (Random Forest (RF) and Adaboost) can be used alongside to predict cowpea yield in Nigeria. This will help the government to plan and will also ensure food security for the people of the state. Overall, the comparative results suggest that while AdaBoost can be useful, Random Forest outperforms it in terms of accuracy and stability. Future work could involve optimizing hyperparameters and exploring hybrid models for enhanced performance.

REFERENCES

Cao, Y., Miao, Q., Liu, J., & Gao, L. (2013). Advance and prospects of AdaBoost algorithm. *Acta Automatica Sinica, 39*, 745-758. https://consensus.app/papers/advance-and-prospects-of-adaboost-algorithm/7a978497301a5d51821174ffb11ffd3c/?utm_source=chatgpt

Charlin, L. (2004). AdaBoost and learning algorithms: An introduction. https://consensus.app/papers/adaboost-and-learning-algorithms-an-introduction-charlin/c895b5e4c8895cbd891176162f1d5ff2/?utm_source=chatgpt

Chandraprabha, & Dhanaraj. (2023). Rice crop yield forecasting using AdaBoost and CNN with Horse Herd Optimization Algorithm. *International Journal of Agricultural Data Science*.

Iliyasu U., Obunadike G.N., & Jiya E. A. (2023) Rainfall Prediction Models for Katsina State, Nigeria: Machine Learning Approach. *International Journal of Science for Global Sustainability*, 9 (2), 151 – 157. DOI: https://doi.org/10.57233/ijsgs.v9i2.473

Jiya, E. A., Illiyasu, U., & Akinyemi, M. (2023). Rice Yield Forecasting: A Comparative Analysis of Multiple Machine Learning Algorithms. *Journal of Information Systems and Informatics,* 5(2), 785-799.

Koduri, et al. (2019). Predictive modeling for crop production using AdaBoost regression. *Journal of Agricultural Informatics*, 10(2), 45-58.

Mahesh, & Soundrapandiyan. (2024). CatBoost: A high-precision approach for agricultural yield prediction. *Machine Learning in Agriculture*, 12(1), 89-103.

Olanrewaju, O. M., Jiya, E. A., & Echobu, F. O. (2024). Intelligent Maize Yield Prediction Model Based on Plant Attributes and Machine Learning Algorithms. *International Journal of Research and Scientific Innovation*, 11(7), 1097-1104.

Pande, R. (2020). Performance evaluation of Random Forest in yield prediction. *International Journal of Data Science and Agriculture*, 8(3), 120-134.

Renju, & Brunda. (2024). Stacking ensemble learning for crop yield prediction. *Journal of AI in Agriculture*, 15(1), 50-65.

Shawon, et al. (2023). A comparative study of deep learning models for precision agriculture. *Computational Agriculture Journal*, 14(2), 98-110.

Shuaibu, et al. (2024). Application of Decision Tree Regressor in Nigerian agriculture. *African Journal of Agricultural Analytics*, 6(2), 75-89.

Mishra, U., Raj, R., & Shukla, D. (2024). Soil Moisture Measurement Using Sensor Based on Cloud IoT. *2024 International Conference on Computing, Sciences and Communications (ICCSC)*, 1-5. https://doi.org/10.1109/ICCSC62048.2024.10830358.

Alsayaydeh, J., Yusof, M., Magenthiran, M., Hamzah, R., Mustaffa, I., & Herawan, S. (2024). Empowering crop cultivation: harnessing internet of things for smart agriculture monitoring. *International Journal of Electrical and Computer Engineering (IJECE)*. https://doi.org/10.11591/ijece.v14i5.pp6023-6035.

Reshma, R., Sathiyavathi, V., Sindhu, T., Selvakumar, K., & Sairamesh, L. (2020). IoT based Classification Techniques for Soil Content Analysis and Crop Yield Prediction. *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 156-160. https://doi.org/10.1109/I-SMAC49090.2020.9243600.